



**University of
Zurich^{UZH}**

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2015

Reproducible and consistent quantification of the *Saccharomyces cerevisiae* proteome by SWATH-mass spectrometry

Selevsek, Nathalie ; Chang, Ching-Yun ; Gillet, Ludovic C ; Navarro, Pedro ; Bernhardt, Oliver M ;
Reiter, Lukas ; Cheng, Lin-Yang ; Vitek, Olga ; Aebersold, Ruedi

Abstract: Targeted mass spectrometry by selected reaction monitoring (S/MRM) has proven to be a suitable technique for the consistent and reproducible quantification of proteins across multiple biological samples and a wide dynamic range. This performance profile is an important prerequisite for systems biology and biomedical research. However, the method is limited to the measurements of a few hundred peptides per LC-MS analysis. Recently, we introduced SWATH-MS, a combination of data independent acquisition and targeted data analysis that vastly extends the number of peptides/proteins quantified per sample, while maintaining the favorable performance profile of S/MRM. Here we applied the SWATH-MS technique to quantify changes over time in a large fraction of the proteome expressed in *Saccharomyces cerevisiae* in response to osmotic stress. We sampled cell cultures in biological triplicates at six time points following the application of osmotic stress and acquired single injection data independent acquisition data sets on a high-resolution 5600 tripleTOF instrument operated in SWATH mode. Proteins were quantified by the targeted extraction and integration of transition signal groups from the SWATH-MS datasets for peptides that are proteotypic for specific yeast proteins. We consistently identified and quantified more than 15,000 peptides and 2500 proteins across the 18 samples. We demonstrate high reproducibility between technical and biological replicates across all time points and protein abundances. In addition, we show that the abundance of hundreds of proteins was significantly regulated upon osmotic shock, and pathway enrichment analysis revealed that the proteins reacting to osmotic shock are mainly involved in the carbohydrate and amino acid metabolism. Overall, this study demonstrates the ability of SWATH-MS to efficiently generate reproducible, consistent, and quantitatively accurate measurements of a large fraction of a proteome across multiple samples.

DOI: <https://doi.org/10.1074/mcp.M113.035550>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-120437>

Journal Article

Accepted Version

Originally published at:

Selevsek, Nathalie; Chang, Ching-Yun; Gillet, Ludovic C; Navarro, Pedro; Bernhardt, Oliver M; Reiter, Lukas; Cheng, Lin-Yang; Vitek, Olga; Aebersold, Ruedi (2015). Reproducible and consistent quantification of the *Saccharomyces cerevisiae* proteome by SWATH-mass spectrometry. *Molecular Cellular Proteomics*, 14(3):739-749.

DOI: <https://doi.org/10.1074/mcp.M113.035550>

Reproducible and consistent quantification of the *Saccharomyces cerevisiae* proteome by SWATH-MS

Authors

Nathalie Selevsek¹, Ching-Yun Chang², Ludovic C Gillet¹, Pedro Navarro¹, Oliver M Bernhardt³, Lukas Reiter³, Lin-Yang Cheng², Olga Vitek⁴, Ruedi Aebersold^{1,6}

Institution

¹Department of Biology, Institute of Molecular Systems Biology, ETH Zürich, 8093 Zürich, Switzerland

²Department of Statistics, Purdue University, West Lafayette IN, USA

³Biognosys, 8952 Zürich, Switzerland

⁴College of Science, Northeastern University, 02115 Boston MA, USA

⁴College of Computer and Information Science, Northeastern University, 02115 Boston MA, USA

⁶Faculty of Science, University of Zürich, Switzerland

Corresponding author:

Prof. Ruedi Aebersold

Institute of Molecular Systems Biology

Wolfgang-Pauli-Str. 16

ETH Zürich

CH-8093 Zürich

Phone: +41 44 633 3170

Fax: +41 44 633 1051

aebersold@imsb.biol.ethz.ch

Running Title

Proteome quantification by SWATH-MS

Keywords

Protein quantification, yeast, targeted, SWATH-MS, DIA

Abstract

Targeted mass spectrometry by selected reaction monitoring (S/MRM) has proven to be a suitable technique for the consistent and reproducible quantification of proteins across multiple biological samples and a wide dynamic range. This performance profile is an important prerequisite for systems biology and biomedical research. However, the method is limited to the measurements of a few hundred peptides per LC-MS analysis. Recently, we introduced SWATH-MS, a combination of data independent acquisition (DIA) and targeted data analysis that vastly extends the number of peptides/proteins quantified per sample, while maintaining the favorable performance profile of S/MRM. Here we applied the SWATH-MS technique to quantify changes over time in a large fraction of the proteome expressed in *Saccharomyces cerevisiae* in response to osmotic stress.

We sampled cell cultures in biological triplicates at six time points following the application of osmotic stress and acquired single injection DIA datasets on a high-resolution 5600 tripleTOF instrument operated in SWATH mode. Proteins were quantified by the targeted extraction and integration of transition signal groups from the SWATH-MS datasets for peptides that are proteotypic for specific yeast proteins. We consistently identified and quantified more than 15,000 peptides and 2500 proteins across the 18 samples. We demonstrate high reproducibility between technical and biological replicates across all time points and protein abundances. In addition, we show that the abundance of hundreds of proteins was significantly regulated upon osmotic shock, and pathway enrichment analysis revealed that the proteins reacting to osmotic shock are mainly involved in the carbohydrate and amino acid metabolism. Overall, this study demonstrates the ability of SWATH-MS to efficiently generate reproducible, consistent and quantitatively accurate measurements of a large fraction of a proteome across multiple samples.

Introduction

In systems biology and biomedical studies targeted mass spectrometry via selected reaction monitoring (SRM) (also known as multiple reaction monitoring, MRM) has emerged as a powerful technique for the consistent and reproducible quantification of proteins across numerous complex samples [1-6]. Optimal sets of precursor/fragment ion pairs, called transitions, uniquely represent a specific peptide. They constitute a definitive mass spectrometric assay for the detection of targeted peptides, and thus the proteins from which they derive, in the complex matrix of trypsinized biological samples [1, 7]. Protein quantification is then performed by relating the intensity of the acquired transition signals to suitable reference signals. Most quantification strategies commonly used in proteomics are compatible with this method [8]. Recently, the high-throughput development of S/MRM assays has been achieved via the generation of MS/MS spectral libraries from the measurements of thousands of synthetic peptides representing proteotypic peptides [9]. Moreover, many experimental and bioinformatics workflows have been developed for assay generation, assay optimization, data evaluation and the dissemination of optimized S/MRM assays [10-16]. In combination, these developments have supported the creation of mass-spectrometric maps of entire proteomes of selected species including *Streptococcus pyogenes*, *Mycobacterium tuberculosis* and *Saccharomyces cerevisiae* [5, 17-19] and the robust use of these resources to quantify specific protein sets across multiple biological samples.

Currently, targeted proteomics by S/MRM can be multiplexed to a maximum set of approximately 100 proteins that can be measured in a single LC-S/MRM run at optimal quantitative accuracy, limit of detection and dynamic range. The quantification of higher numbers of proteins per run compromises some of the performance parameters of the method due to well understood tradeoffs [8]. Attempts have been made to further increase the degree of multiplexing of S/MRM, either by automated adjustment of the scheduled detection windows [20] or by acquiring, in a data-dependent manner, the complete set of precursor-fragment ion pairs of a given assay [21]. Alternatively, parallel reaction monitoring (PRM) approach operated on quadrupole-orbitrap mass spectrometer has shown detection

and quantification performances similar or better than those obtained in SRM, due to the increased selectivity of the mass analyzer [22-24]. These approaches are promising, but their application relies on prior knowledge of the precursor ions that need to be targeted during the data acquisition, and they still are subject of the above mentioned tradeoffs.

Recently, we developed a novel MS strategy that combines data independent acquisition (DIA) of trypsinized protein samples with S/MRM-like, *in silico* targeted analysis of the acquired complete fragment ion maps [25]. We termed the method SWATH-MS, and applied the sequential isolation window acquisition principle [26] to repeatedly cycle, in a single injection, through 32 consecutive 25-Da precursor isolation windows (swaths). The process acquires fragment ion spectra of all precursors in a space defined by the 400-1200 m/z precursor range and a user-specified retention time window. We used the prior information in MS/MS spectral libraries to extract groups of signals that uniquely identify a specific peptide, and to demonstrate that peptides could be identified and quantified over a dynamic range of 4 orders of magnitude, even when the precursors were not detectable in a survey MS scan. For the 45 proteins involved in the central carbon metabolism of yeast we demonstrated that the accuracy of quantification was equivalent to that of S/MRM [25]. However, due to the lack of adequate software tools at that time, the extensive high-throughput targeted data analysis of the SWATH-MS maps could not be fully demonstrated in that first study.

Here we demonstrate the multiplexing capabilities of SWATH-MS for the detection and quantification of significantly larger fractions of a proteome as compared to S/MRM, without compromising reproducibility, consistency and quantitative accuracy. We describe the large scale deployment of fragment ion spectral libraries and the use of S/MRM-like analysis tools specifically adapted to SWATH-MS data for the detection and quantification of temporal changes of the *S. cerevisiae* proteome in response to osmotic stress.

Material and Methods

Yeast culture, protein isolation and digestion

Three series of six cultures each from the yeast strain BY4741 MATa his3 Δ leu2 Δ met15 Δ ura3 Δ were grown in SD medium until they reached an A₆₀₀ of 0.8. To apply the osmotic shock, 0.4 M NaCl was added to each 50 ml culture and after 0 min (T0), 15 min (T1), 30 min (T2), 60 min (T3), 90 (T4) and 120 min (T5) the cells were harvested. At the respective time points the culture media were quenched by addition of trichloroacetic acid (TCA) to a final concentration of 6.25 % and the cells were harvested by centrifugation at 1500 g for 5 min at 4°C. The supernatants were discarded and the cell pellets were washed three times by centrifugation with cold (-20°C) acetone to remove interfering compounds. The final cell pellets were resolubilized in lysis buffer containing 8 M urea, 0.1 M NH₄HCO₃ and 5 mM EDTA and cells were disrupted by glass bead beating (5 times 5 minutes at 4°C). The total protein amount from the pooled supernatants was determined by BCA Protein Assay Kit (Thermo, Rockford, US). Yeast proteins were reduced with 12 mM dithiotreitol at 37°C for 30 min and alkylated with 40 mM iodoacetamide at room temperature in the dark for 30 min. Samples were diluted with 0.1 M NH₄HCO₃ to a final concentration of 1.5 M urea and the proteins were digested with sequencing grade porcine trypsin (Promega) at a final enzyme:substrate ratio of 1:100. Digestion was stopped by adding formic acid to a final concentration of 1%. Peptide mixtures were desalted using reverse phase cartridges Sep-Pak tC18 (Waters, Milford, MA) according to the following procedure; wet cartridge with 1 volume of 100% methanol, wash with 1 volume of 80% acetonitrile, equilibrate with 4 volumes of 0.1% formic acid, load acidified digest, wash 6 volumes of 0.1% formic acid, and elute with 1 volume of 50% acetonitrile in 0.1% formic acid. Peptides were dried using a vacuum centrifuge and resolubilized in 100 μ l of 0.1% formic acid and frozen at -20°C. For in depth-fractionation experiments of NaCl-untreated yeast cells, the peptide mixtures were separated by off gel electrophoresis (OGE) using a pH 3–7 IPG strip (Amersham Biosciences) and a 3100 OFFGEL Fractionator (Agilent Technologies) with collection in 12 wells and then submitted for C18 clean-up. All samples were spiked with the retention time standard peptides iRT-Kit (Biognosys, Schlieren, Switzerland).

LC-MS set up

Yeast tryptic peptides were analyzed on a 5600 TripleTof™ mass spectrometer (ABSciex, Concord, Ontario, Canada). Chromatographic separation of peptides was performed on a NanoLC-2Dplus HPLC system (Eksigent, Dublin, CA) coupled to a 75 µm diameter fused silica emitter packed with 20 cm of Magic C18 AQ 3 µm resin (Michrom BioResources, Auburn, CA, USA). Peptide samples (2 µl injections containing 2 µg of total peptide amount) were loaded on the column from a cooled (4°C) Eksigent autosampler using a flow of 300 nl/min of an isocratic 98% Buffer A (Buffer A: 2% acetonitrile, 0.1% formic acid) and 2% Buffer B (Buffer B: 98% acetonitrile, 0.1% formic acid). Peptides were then separated at a flow rate of 300 nl/min with a 180 minutes linear gradient of 2% to 35% Buffer B. For shotgun experiments the mass spectrometer was operated with a “top20” method: Initially, a 500 ms survey scan (TOF-MS) was collected from which the top 20 precursor ions were automatically selected for fragmentation, whereby each MS/MS event consisted of a 150 ms fragment ion scan. The main selection criterion for parent ion selection was precursor ion intensity. Ions with an intensity exceeding 200 counts per second and a charge state greater than 2+ were preferentially selected. The selected precursors were then added to a dynamic exclusion list for 20s. Ions were isolated using a quadrupole resolution of 0.7 amu and fragmented in the collision cell using the collision energy equation ($0.0625 \times m/z - 3.5$) with an additional collision energy spread of 15 eV within the 150 ms accumulation time to mimic SWATH like fragmentation conditions. If ≤ 20 precursor ions meeting the selection criteria were detected per survey scan, the detected precursors were subjected to extended MS/MS accumulation times to maintain a constant total cycle time of 3.5s. For SWATH-MS based experiments the mass spectrometer was operated in a looped product ion mode. In this mode the instrument was specifically tuned to allow a quadrupole resolution of 25 amu per mass selection. The stability of the mass selection was maintained by the operation of the RF and DC voltages on the isolation quadrupole in an independent manner. Using an isolation width of 25 amu, a set of 32 overlapping windows (1 amu overlap) was constructed covering the mass range 400 to 1200 amu. The collision energy for each window was determined based on the collision energy for a putative 2+ ion centered in the respective window (equation: $0.0625 \times m/z - 3.5$) with a collision energy spread of 15 eV. An accumulation time of 100 ms was used for each fragment ion scan and for

the survey scans acquired at the beginning of each cycle, resulting in a total cycle time of 3.3 s. The sequential precursor isolation window set-up was as follows: 499-425, 424-450, 449-475, ... 1174-1200, with an effective (100%) transmission of ~25 Da and ~0.3 Da attrition on either side of the isolation window.

Generation of spectral library and database searching

The shotgun spectral library was generated using a total of 46 shotgun injections on an ABSciex TripleTOF 5600 instrument: 2 series (technical replicate injections) of 12 off-gel fractions from the BY 4741 strain (in exponential phase, OD600 = 0.8), 18 injections (6 time points in biological triplicates) from the complete BY 4741 osmotic shock study, and 4 technical replicate injections of a FY4 strain pooled from various time points of a diauxic shift experiment [1]. Each file was searched with Sorcerer-Sequest (TurboSequest v4.0.3rev11 running on a Sage-N Sorcerer v4.0.4) and Mascot (version 2.3.0) against the SGD database (release 03 Febr. 2011, containing 6750 yeast protein entries, concatenated with 6750 corresponding “tryptic peptide pseudo-reverse” decoy protein sequences). For the search, we allowed for semi-tryptic digests and up to 2 missed cleavages per peptide, and we used carbamidomethylation as a fixed modification on cysteine and oxidation as variable modification on methionine residues. The Sequest and Mascot search results were converted to pep.xml and then combined using iProphet (included in TPP version 4.5.2). The search results were sorted by decreasing iProphet probability and filtered at 1% FDR by decoy counting (iProphet score cut-off 0.0242) at the peptide spectrum matches (PSM) level, resulting in 891’570 identified spectra, 78’605 unique peptides (4.7% FDR at the peptide level) and 5’125 proteins (inclusive isoforms). Those spectra were used to build a consensus spectra library using SpectraST (included in TPP 4.5.2) [27]. The transition MS/MS coordinates for the peptides were then computed using an in-house python script that used the spectrast .sptxt library as input and retrieved the top 3-4 most intense (singly or doubly charged) y or b fragment ions for each spectra applying the following algorithm: i) the fragment ion m/z was above 300 and outside of the range of the 25 Da swath/precursor fragmentation window from which the fragment ion was acquired and ii) the extracted m/z values matched the theoretical fragment ion

masses within 0.05 Da tolerance. We only selected the transitions originating from proteotypic peptides (i.e. those uniquely matching to a single protein isoform from the SGD database) and that did not contain oxidation on methionine. The final fragment ion library comprised 331'449 transitions for 83'520 proteotypic precursors (66'007 unique modified peptides matching to 4'596 unique protein isoforms). For each peptide, we appended its iRT value by calculating the average of the iRT values found for the corresponding precursor(s) of that peptide at 1% PSM FDR in the search results across all the shotgun runs by using a simple linear regression after re-alignment/re-scaling onto the spiked-in reference iRT peptides (Biognosys, Schlieren, Switzerland).

SWATH- MS targeted data extraction

Targeted data extraction of the SWATH-MS data was performed by the *Spectronaut* software from Biognosys (RC.2.0.3). Spectronaut processed the SWATH data using the above-mentioned assay list of target peptides and using a targeting extraction and scoring strategy similar to that of the S/MS analysis tool *mProphet* [15]. In addition to the subscores used by *mProphet*, *Spectronaut* also used retention time prediction based on iRT [28], the m/z dimension in the SWATH MS data, mass accuracy and isotopic distribution of fragment ions to identify a peptide. A maximum of four transitions was extracted for each targeted peptide, together with their corresponding decoy-transition groups, which were generated by pseudo-reversing the sequence of the targeted peptides.

False discovery rates (FDR) were determined for SWATH-MS using an adapted decoy model similar to that used by *mProphet* [15]. This method is based on three critical steps, which can be used in any targeted proteomics experiment. The first step involves the generation of signal groups for decoy analytes, such that the resulting signals are consistent with the patterns in the real data, but correspond to undetectable analytes (Figure S1). We assessed the quality of this step by the following analyses: we targeted three classes of peptides in a yeast sample, i) yeast peptides likely detectable in the sample (e.g. as determined by their identification by shotgun analysis of the sample), ii) decoy peptides generated from the yeast peptides, and iii) peptides that are not present in the sample, e.g. peptides from human proteins which are not contained in yeast proteins. If the identified decoy peptides are

truly representing false identifications, the resulting (noise) signals have to be similar and the resulting score distributions have to overlap. In supplementary Figure S2A we performed such an analysis, and showed that the employed decoy model very accurately represented false identifications. The second step involves fitting a representative probability distribution to the histogram of the decoy scores. Although in principle various types of probability distributions could be fit, we observed that with the data at hand a Gaussian distribution provides a very good fit (supplementary figure S2A). The fitted distributions allow us to calculate, separately for each peptide transition, the p-value of incorrect identification. As is well known, p-values do not account for the multiplicity of the peptide in the experiment. However, multiple methods allow us to use the p-values to control the FDR in the list of the identifications. We used the method of Storey et al. [29], which uses a two-group mixture model of the p-values to estimate the q-values, and to control the FDR. In this study, the estimates based on this method turned out to be very accurate and generally slightly conservative (supplementary Figure S2B). This process of FDR estimation in targeted proteomics experiments has also been validated in a similar manner for S/MRM. For a detailed description please see the supplementary of [15].

For the technical replicate measurements, the features confidently identified below 1% FDR threshold were used to estimate the elution iRT value of the feature. Then we recovered the features with q-values over 1% FDR but below 10% FDR, for which an elution iRT value was close (i.e., less than 0.35 iRT units) to the estimated elution iRT. To evaluate the FDR after this identification recovery, we considered that the iRT tolerance window only covers 8.33% of the total extraction window, and therefore only 8.33% of the recovered false identifications meet the recovery criterion, assuming that false identifications are spread uniformly in the retention time space. This keeps the actual FDR below 1%.

Statistical analysis The investigation aimed at identifying proteins that significantly changed their abundance across the different time points. All the identified and quantified peak intensities were first transformed by the logarithm with base 2. To further filter out incorrectly identified or quantified transitions, we calculated the Pearson correlation between the intensity of each transition and the

average intensity of the protein's transitions across all MS runs, and transitions with correlation below 0.5 were removed. The remaining data were subjected to constant normalization to equalize the median peak intensities of transitions between runs [30].

Protein-level quantification and testing for differential abundance were performed using MSstats[31] based on a linear mixed-effects model [32]. The model decomposes log-intensities into the effects of times, biological replicates, SWATH features (i.e., a combination of peptides and transitions), MS runs, and statistical interactions. The model specified the reduced scope of biological replication. For each investigated comparison, MSstats provided model-based estimates of fold changes, as well as p-values that were adjusted to control the False Discovery Rate (FDR) at the cut-off 0.05 [33]. All the input and output files from the MSstats analysis together with the R scripts performed for the manuscript are provided in Supplementary data 1.

Results

Detection of yeast proteins by SWATH-MS

We first determined the fraction of the *S. cerevisiae* proteome that could be detected in a trypsinized, unfractionated yeast cell lysate by SWATH-MS. For that purpose, we selected MS coordinates consisting of precursor/ fragment ion pairs, the relative fragment ion intensities and peptide elution time for each targeted peptide, and extracted signal groups corresponding to these coordinates from SWATH-MS datasets. The MS coordinates were generated as prior information via in-depth fractionation and MS sequencing of yeast cell digests using the same model 5600 QqTOF mass spectrometer that was also used to acquire the SWATH-MS datasets (Figure 1). For the DDA measurements we generated protein extracts from cells in different perturbed conditions, including stages of a diauxic shift experiment and cells subjected to salt stress. Tryptic digests of these samples were fractionated by isoelectric focusing via off-gel electrophoresis and analyzed by DDA MS. The resulting MS/MS spectra were searched against the *Saccharomyces* Genome Database (SGD) and the identified peptides were statistically filtered at 1% FDR at the PSM level. We compiled the resulting MS/MS spectra into a spectral library using SpectraST [27] and reported the final number of spectra,

peptides and proteins in supplementary Results Figure S3 and Table S1. Next, we used this spectral and an in-house python script to extract the coordinates of MS/MS transitions to target in SWATH-MS (supplementary Table S2). In total, we targeted 331'449 fragment ion traces (transition signals) corresponding to 66'007 proteotypic peptides (PTPs) and 4'596 proteins in each SWATH-MS. The fragment ion traces were extracted using the *Spectronaut* software, and the results were filtered at 1% FDR at the assay level using the error rate model originally described for *mProphet* [15]. The results indicated that 16178 peptides corresponding to 2578 proteins could be confidently detected with a FDR of 1 % estimated by *mProphet* in an unfractionated digest of a yeast cell lysate (supplementary Table S3). Overall, 71 % of the proteins were identified with at least 2 peptides per protein (Figure 2A). According to the comprehensive western blot analysis performed by tandem affinity purification tagged yeast ORFs [34] the detected proteins spanned a concentration range of 4 orders of magnitude down to $1e^6$ to 100 copies/cell (figure 2B). Furthermore, the data showed that the detected proteins were not biased against low abundant proteins, and that SWATH-MS could confidently detect > 300 proteins that could not be detected by the quantitative western blot analysis (figure 2C). The fraction of the spectral library that was not detected by SWATH-MS may be explained by the fact that SWATH-MS is still not sensitive at the level of the SRM technology [35]. Overall, these results demonstrate that a single-injection data independent acquisition (DIA), combined with targeted data extraction, has the power to detect more than 2500 proteins spanning the dynamic range of protein expression in yeast down to 100 copies/cell.

Reproducibility and consistency of proteome measurements by SWATH-MS

We next tested whether the same fraction of the yeast proteome could be consistently and reproducibly detected and quantified across multiple MS injections. For this purpose, we acquired several (n= 4) SWATH-MS datasets from an unfractionated tryptic digest of a yeast sample composed of extracts of cells at different states, and used the *Spectronaut/ mProphet* software tools for targeted data extraction of the queried proteins. In total we detected an average of 18600 \pm 72 unique peptides corresponding to 2880 \pm 7 proteins per single run with an estimated FDR of 1 % at the assay level (supplementary

Table S4). Overall we identified on average 28% of peptides (18600 / 66007) present in the spectral library, where 80% of those could be confidently detected in all four injections, and more than 90% in three of four injections (Figure 3A). To evaluate the reproducibility of the method for proteome wide quantification, we determined the coefficients of variation (CV%) of the integrated transition peak areas across injections. We integrated the fragment ions traces from assays that were confidently detected in all 4 runs, which corresponded to a total of 17552 peptides and 2333 proteins, respectively. Figure 3B indicates that for the majority of the assays (76 %) the observed CV was ≤ 10 %, and that for a further 20% of assays the CVs were between 10 and 40%. Furthermore, the results showed that assays detected with high signal intensities (>500) were quantified with CV's comparable to those detected with lower signal intensities (<100) (Figure 3C), demonstrating that peptides can be reproducibly quantified over a dynamic range of 4 orders of magnitude by SWATH-MS. Overall, these data indicate that the method has the capability of detecting a significant fraction of the proteome across multiple injections at a high degree of reproducibility, and to quantify proteins with high consistency across minimally four orders of magnitude dynamic range.

Accuracy of proteome quantification by SWATH-MS

To evaluate the accuracy of quantification of large sets of proteins targeted in SWATH-MS datasets we prepared two mixtures containing tryptic digests of yeast cultures grown either on ^{14}N - or ^{15}N -labeled ammonium sulfate medium and mixed the two samples in different ratios. The first mixture was a yeast digest containing equal amounts of ^{14}N - and ^{15}N -labeled proteins (i.e. 1:1; $^{14}\text{N} : ^{15}\text{N}$). The second mixture contained ^{14}N -labeled proteins at a tenfold excess over the ^{15}N -labeled proteins. (i.e. 10:1; $^{14}\text{N} : ^{15}\text{N}$). Each of the mixtures was prepared in triplicates and measured in SWATH mode. Signals for the queried proteins were extracted with the *Spectronaut/mProphet* software and analyzed as described above. We integrated the fragment ions traces from those peptides that were confidently identified in all 6 runs and in both isotopic channels, i.e. a total of 3354 peptides from 780 proteins. In the case of the 1: 1 mixture, an estimated fold change of 0.92 ± 0.14 SE was obtained whereas for the 1: 1/10 mixture, 10.47 ± 4 SE by MSstats (Figure 4), thus demonstrating that SWATH-MS achieves

accurate relative quantification at the proteome level. The quantification results of the 6 SWATH-MS runs are represented in supplementary Table S5.

Quantification of changes in the yeast proteome upon osmotic shock

We next deployed the technique to quantify the changes in the *S. cerevisiae* proteome induced by osmolarity stress. The osmolarity stress response has been extensively studied in yeast and it occurs frequently in the yeast cell's natural environment. Specifically, we investigated the time-resolved response of the yeast proteome to NaCl, a salt that is commonly used for inducing an osmotic stress response [36]. We added NaCl to a concentration of 0.4 M to cells in exponential growth phase and sampled the cultures at different time points (0 min (T0), 15 min (T1), 30 min (T2), 60 min (T3), 90 min (T4) and 120 min (T5)) after salt addition in biological triplicates. The proteins isolated from the 18 yeast cultures were digested and subjected to SWATH-MS analysis. Targeted data extraction of the queried proteins was performed using the spectral library described above. The results of the targeted data extraction are presented in supplementary Table S6. Identified and quantified peak intensities were log-transformed and normalized using MSstats as described in material and methods. For the downstream statistical analysis we only considered the peptides that were detected in at least 3 replicates and at least one time point. As the result of this filtering, 29 % of transitions were removed before the statistical analysis, but after the identification step. In other words, the protein identification step by Spectronaut used four transitions for every protein, while the MSstats analysis step used between one and four transitions per protein. After all the filtering steps, the input to MSstats was a set of 2589 target proteins, where each protein was represented by 1–129 peptides, and each peptide was represented by 1–4 transitions and no missing values. Out of the 2589 proteins, only 75 proteins had one transition as input to MSstats.

Next, protein-level quantification and testing for differential abundance were performed based on a linear mixed-effects model [32]. Several comparisons with respect to the baseline at time point 0 min (T0) (specifically, comparing protein abundances at time 0 min (T0) to the abundances at 15 min (T1),

30 min (T2), 60 min (T3), 90 min (T4) and 120 min (T5) after addition of salt) were tested for each protein.

Next, we performed unsupervised clustering of proteins according to their patterns of change in abundance over time. We enumerated all the possible 3^5 such patterns (i.e. significantly up-regulated, significantly down-regulated and no statistically significant change, at times 15 min (T1), 30 min (T2), 60 min (T3), 90 min (T4) and 120 min (T5) as compared to time 0 min (T0)). We retained four clusters with more than 50 proteins each for further examination (Figures 5A and 5B). Proteins in cluster 1 ($n = 266$) and cluster 2 ($n = 67$) are up-regulated along the complete time course, with a > 20 min delay in the response for proteins belonging to cluster 2. Pathway enrichment analysis tool DAVID (<http://david.abcc.ncifcrf.gov/>) revealed an over-representation of proteins involved in carbohydrate metabolism such as the glycolysis-gluconeogenesis pathway ($p = 4.4e^{-6}$), the starch and sucrose metabolism ($p = 7.2e^{-4}$) and the pentose phosphate pathway ($p = 8.3e^{-3}$). These pathways are directly linked to the glycerol, trehalose and glycogen metabolism, which are known to be induced by osmotic shock to trigger the production of glycerol, an essential osmolyte for osmoadaptation [36]. In contrast, downregulated protein profiles were obtained for cluster 3 ($n = 219$) and cluster 4 ($n = 567$) upon addition of salt, with a > 20 min delay in the response for proteins belonging to cluster 3. Mainly enzymes involved in amino acid biosynthesis (i.e. the glycine, serine, threonine metabolism pathway; $p = 3e^{-4}$ and the phenylalanine, tyrosine and tryptophan pathway, $p = 2.3e^{-2}$) were found to be repressed as previously suggested [37].

Validation of protein fold changes by S/MRM

To validate the protein fold changes obtained by SWATH-MS, we quantified a subset of 100 proteins in the 18 osmotic shock time course experiment using S/MRM. Among these proteins, 24 were up-regulated along the complete time course according to SWATH-MS, 22 were down-regulated along the complete time course, and 5 were down-regulated along the complete time course after a > 20 min delay. The remaining 49 proteins were not regulated between any time points according to SWATH-MS. For each of the 100 proteins, we chose the highest-responding peptide for each protein that was

confidently identified in SWATH-MS, together with their corresponding fragment ions (i.e. four transitions in total), for S/MRM quantification. The fragment ion traces of these peptides detected with S/MRM were extracted and integrated using Skyline [13] together with the corresponding peptides detected with SWATH-MS (Supplementary data 2). This allowed us to consistently process both data sets with the same integration/quantification parameters using the same software. Since the optimal peptides and transitions were chosen for the S/MRM quantification, all the quantified transitions were of a relatively good quality, and no downstream filtering was necessary. The quantitative values were used as input to MSstats to estimate log fold changes and their associated confidence intervals, and to test proteins for differential abundance, between 15 min (T1), 30 min (T2), 60 min (T3), 90 min (T4), 120 min (T5) and the initial time point 0 min (T0).

Supplementary data 3 shows log fold change profiles and their associated confidence intervals of the 51 regulated proteins quantified by S/MRM and SWATH-MS. To be consistent, for the results for SWATH-MS we only used the subset of the peptides and of the fragments that were also quantified by S/MRM, even though for SWATH-MS additional transition signals for these peptides and additional peptides were concurrently acquired. The data show that the majority of the confidence intervals overlap, indicating good agreement between SRM and SWATH-MS quantification. To further formalize the comparison between the methods, we compared the outcomes of tests for differential abundance. We classified the outcome of each test for differential abundance as significant up-regulation (denoted as 1), significant down-regulation (denoted as -1), and absence of significant regulation (denoted as 0) (supplementary Table 7). These tests for differential abundance were applied to the 5 time points of the study resulting in five values per protein across the time course. Table 1 shows that for 64.7% of the proteins the conclusions from the two datasets agree in at least 4 time points. Conversely, only for 17.6% of the proteins conclusions from the two datasets agree in less than 3 time points, whereas 7 proteins had no agreement due to interferences or low quality peak shape. (see supplementary data 4). The remaining discrepancies can also be due to other reasons. First, the S/MRM and SWATH-MS acquisitions occurred several weeks apart. Second, external factors such differences in chromatographic conditions (3 hours in SWATH versus. 30 min in SRM) may generate

differences in ionisation or ion suppression between both measurements, in a way that impacts peptide quantification accuracy of one or the other method. Finally, the nature of the mass analysis between the two methods may result in differences in sensitivity and detection. Overall, the results demonstrated that SWATH-MS targeted analysis of complex samples can provide biological conclusions that are in a high agreement with S/MRM, but at a much higher throughput.

Correlation of protein fold changes with transcriptomics

We next correlated the protein abundance profiles obtained by SWATH-MS with their corresponding transcript profiles across the four most significantly regulated pathways (i.e. the glycolysis-gluconeogenesis pathway, the pentose phosphate pathway, the glycine, serine, threonine metabolism pathway and the phenylalanine, tyrosine and tryptophan pathway). For this purpose, we used a transcriptomic data set that was previously generated for yeast treated under similar experimental conditions [37]. Figure 6 represents the quantitative patterns (as heat maps) for each transcript and protein along the osmotic shock time course together with a frequency plot assessing of their similarity in terms of Pearson correlation coefficients for each of the four pathways. The results revealed that the correlation between transcripts and protein abundance greatly differs depending on the biological pathway. In the case of the glycolysis-gluconeogenesis pathway, 50 % of the detected proteins are reflecting good levels of correlation (i.e. Pearson correlation coefficient > 0.5) with their corresponding transcripts, whereas 50 % of the detected proteins involved in the phenylalanine, tyrosine and tryptophan pathway are showing consistent anti-correlation (i.e. Pearson correlation coefficient < 0.5) between protein and transcript abundance changes. However, 50 % of the protein profiles measured for the pentose phosphate pathway and the glycine, serine and threonine metabolism pathway showed no clear correlation (i.e. $-0.5 < \text{Pearson correlation coefficient} < 0.5$) between the protein abundance and their corresponding RNA profiles, which may be mainly due to a slight delay observed for the protein response compared to the RNA response. Protein and transcript profiles for all genes can be found in the in Supplementary data 5.

Discussion

Reliable and precise quantitative data sets can be acquired through the consistent recording of the MS signals of a peptide across all the desired conditions or perturbations in a biological system. Currently, two MS strategies are fulfilling such criteria, namely S/MS and “data independent acquisition” (DIA). In S/MS, the instrument acts as a mass filter and selectively target only the precursor and fragment ions of interest. The resulting chromatographic peak groups are then analysed by using the information (i.e. fragmentation pattern or chromatographic properties) available in MS/MS spectral libraries. In the case of DIA-MS, MS/MS spectra are triggered within a user defined ion window independently of precursor intensities and then analysed with traditional database search algorithms.

Recently we developed the SWATH-MS technology that combines the DIA acquisition approach with targeted data analysis of S/MS [35] and could thus demonstrate on a set of 60 peptides higher sensitivity than analysing the data with regular database searching. In the present study we demonstrated the multiplexing capabilities of SWATH-MS for large scale quantitative proteomics studies. We were able to confidently detect around 2500 proteins spanning the dynamic range of protein expression in yeast in a 3-hours single sample injection. In contrast, S/MS would have required 48 hours of instrument time to detect the same number of proteins in one single sample due to its lower multiplexing capacity (i.e. ~100 proteins/run). Thus, the SWATH-MS measurements for the 18 osmotic shock time course samples were completed in ≤ 3 days. Since the data structure of SWATH-MS data is equivalent to that of S/MS, a similar/extended bioinformatics workflow was implemented. S/MS, several chromatographic peak groups extracted for the same targeted peptide, evaluated using automated and objective probabilistic scoring model. To confidently identify the targeted peptides by SWATH-MS, we applied *Spectronaut*, a bioinformatics tool the scoring strategy *mProphet* was developed for the automatic evaluation of S/MS signals [15]. In addition to the chromatographic (S/MS-like) scores (i.e. coelution, peak shape similarity, intensity and correlation

of fragmentation pattern between peak groups and assays), SWATH-MS specific scores were added such as mass accuracy and isotopic distribution of fragment ions. A combined score (i.e. discriminated score) was then calculated for each detected peak group and finally used for FDR estimation. Thus for the automated analysis of the 18 SWATH-MS runs by *Spectronaut* less ≤ 2 days were required and allowed the detection of ≥ 2500 yeast proteins with high confidence along the time course study. To further pinpoint the proteins that were significantly changing in abundance between the different time points, we applied MSstats, a statistical modeling framework for protein significance analysis previously designed for S/MSRM experiments [32]. MSstats uses an intensity-based approach decomposing the MS signals obtained for each protein across isotopic labels, peptides, charge states, transitions, samples, and conditions. It has been shown that MSstats performed better in terms of sensitivity and accuracy than simple statistical methods like *t* tests. By applying MSstats to our SWATH-MS data sets, we could show that out of the 2589 quantified proteins, 333 and 786 were found significantly up-and down-regulated, respectively, along the complete time course study or with a delay of 20 min. Many of these proteins were shown to be involved in metabolic pathways and were known to be active upon osmotic shock [37]. Besides, we could show on a set of 51 yeast proteins that SWATH-MS delivered similar quantitative protein profiles than S/MSRM along the 18 osmotic shock time course samples. Furthermore, the reproducibility of SWATH-MS runs was demonstrated with low within-run coefficients of variation (CVs) of $\leq 10\%$ for the majority of the targeted peptides, with only minimal dependency of peptide abundance, results that are similar to those obtained by S/MSRM [38]. In conclusion, the results demonstrate that SWATH-MS allows for the quantification of large set of proteins across multiple samples with a precision, reproducibility and accuracy that is comparable to that obtainable by S/MSRM. With the adaptation of S/MSRM-based software tools for the SWATH-MS targeted data analysis, the technique can be rapidly applied to any type of system biology or biomedical investigation, as it was successfully demonstrated in the last years with S/MSRM but with a higher throughput and higher degree of multiplexing.

Acknowledgments

We thanks Lucia Espona for data deposition ; Stefania Vaga and Jonas Grossmann for helpful comments on the manuscript. This work was supported by EU FP7 Unicellsys (Grant No. 201142); ERC Proteomics v3.0 (Grant No. 233226) and ETH Zurich.

Accession codes.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org/>) via the PRIDE partner repository with the data set identifier PXD001010.

The SRM data have been deposited to the PASSEL [11] via the peptide atlas website with the accession number PASS00504.

References

1. Picotti, P., et al., *Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics*. Cell, 2009. **138**(4): p. 795-806.
2. Costenoble, R., et al., *Comprehensive quantitative analysis of central carbon and amino-acid metabolism in Saccharomyces cerevisiae under multiple conditions by targeted proteomics*. Mol Syst Biol, 2011. **7**: p. 464.
3. Lange, V., et al., *Selected reaction monitoring for quantitative proteomics: a tutorial*. Mol Syst Biol, 2008. **4**: p. 222.
4. Addona, T.A., et al., *A pipeline that integrates the discovery and verification of plasma protein biomarkers reveals candidate markers for cardiovascular disease*. Nature biotechnology, 2011.
5. Picotti, P., et al., *A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis*. Nature, 2013. **494**(7436): p. 266-70.
6. Picotti, P. and R. Aebersold, *Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions*. Nat Methods, 2012. **9**(6): p. 555-66.
7. Stahl-Zeng, J., et al., *High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites*. Mol Cell Proteomics, 2007. **6**(10): p. 1809-17.
8. Domon, B. and R. Aebersold, *Options and considerations when selecting a quantitative proteomics strategy*. Nat Biotechnol, 2010. **28**(7): p. 710-21.
9. Picotti, P., et al., *High-throughput generation of selected reaction-monitoring assays for proteins and proteomes*. Nat Methods, 2010. **7**(1): p. 43-6.
10. Deutsch, E.W., et al., *TraML--a standard format for exchange of selected reaction monitoring transition lists*. Mol Cell Proteomics, 2012. **11**(4): p. R111 015040.
11. Farrah, T., et al., *PASSEL: the PeptideAtlas SRMexperiment library*. Proteomics, 2012. **12**(8): p. 1170-5.
12. Brusniak, M.Y., et al., *ATAQS: A computational software tool for high throughput transition optimization and validation for selected reaction monitoring mass spectrometry*. BMC Bioinformatics, 2011. **12**: p. 78.
13. MacLean, B., et al., *Skyline: an open source document editor for creating and analyzing targeted proteomics experiments*. Bioinformatics, 2010. **26**(7): p. 966-8.
14. Abbatiello, S.E., et al., *Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry*. Clin Chem, 2010. **56**(2): p. 291-305.
15. Reiter, L., et al., *mProphet: automated data processing and statistical validation for large-scale SRM experiments*. Nat Methods, 2011. **8**(5): p. 430-5.
16. Sherman, J., et al., *Unique ion signature mass spectrometry, a deterministic method to assign peptide identity*. Mol Cell Proteomics, 2009. **8**(9): p. 2051-62.
17. Picotti, P., et al., *A database of mass spectrometric assays for the yeast proteome*. Nat Methods, 2008. **5**(11): p. 913-4.
18. Karlsson, C., et al., *Proteome-wide selected reaction monitoring assays for the human pathogen Streptococcus pyogenes*. Nat Commun, 2012. **3**: p. 1301.
19. Schubert, O.T., et al., *The Mtb proteome library: a resource of assays to quantify the complete proteome of Mycobacterium tuberculosis*. Cell Host Microbe, 2013. **13**(5): p. 602-12.
20. Gallien, S., et al., *Highly multiplexed targeted proteomics using precise control of peptide retention time*. Proteomics, 2012. **12**(8): p. 1122-33.
21. Kiyonami, R., et al., *Increased selectivity, analytical precision, and throughput in targeted proteomics*. Mol Cell Proteomics, 2011. **10**(2): p. M110 002931.

22. Gallien, S., et al., *Selectivity of LC-MS/MS analysis: implication for proteomics experiments*. J Proteomics, 2013. **81**: p. 148-58.
23. Gallien, S., et al., *Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer*. Mol Cell Proteomics, 2012. **11**(12): p. 1709-23.
24. Peterson, A.C., et al., *Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics*. Mol Cell Proteomics, 2012. **11**(11): p. 1475-88.
25. Gillet, L.C., et al., *Targeted data extraction of the MS/MS spectra generated by data independent acquisition: a new concept for consistent and accurate proteome analysis*. Mol Cell Proteomics, 2012.
26. Venable, J.D., et al., *Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra*. Nat Methods, 2004. **1**(1): p. 39-45.
27. Lam, H., et al., *Building consensus spectral libraries for peptide identification in proteomics*. Nat Methods, 2008. **5**(10): p. 873-5.
28. Escher, C., et al., *Using iRT, a normalized retention time for more targeted measurement of peptides*. Proteomics, 2012. **12**(8): p. 1111-21.
29. Storey, J.D. and R. Tibshirani, *Statistical significance for genomewide studies*. Proc Natl Acad Sci U S A, 2003. **100**(16): p. 9440-5.
30. Zien, A., et al., *Centralization: a new method for the normalization of gene expression data*. Bioinformatics, 2001. **17 Suppl 1**: p. S323-31.
31. Choi, M., et al., *MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments*. Bioinformatics, 2014.
32. Chang, C.Y., et al., *Protein significance analysis in selected reaction monitoring (SRM) measurements*. Mol Cell Proteomics, 2012. **11**(4): p. M111 014662.
33. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: A practical and powerful approach to multiple testing*. J. Roy. Statistical Soc., 1995. **57**: p. 289-300.
34. Ghaemmighami, S., et al., *Global analysis of protein expression in yeast*. Nature, 2003. **425**(6959): p. 737-41.
35. Gillet, L.C., et al., *Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis*. Mol Cell Proteomics, 2012. **11**(6): p. O111 016717.
36. Rep, M., et al., *The transcriptional response of Saccharomyces cerevisiae to osmotic shock. Hot1p and Msn2p/Msn4p are required for the induction of subsets of high osmolarity glycerol pathway-dependent genes*. J Biol Chem, 2000. **275**(12): p. 8290-300.
37. Lee, M.V., et al., *A dynamic model of proteome changes reveals new roles for transcript alteration in yeast*. Mol Syst Biol, 2011. **7**: p. 514.
38. Addona, T.A., et al., *Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma*. Nat Biotechnol, 2009. **27**(7): p. 633-41.

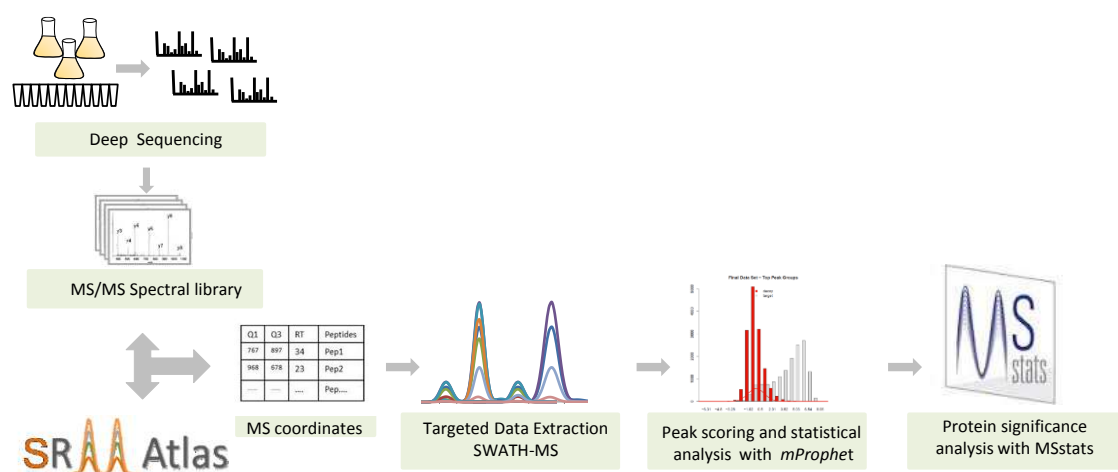
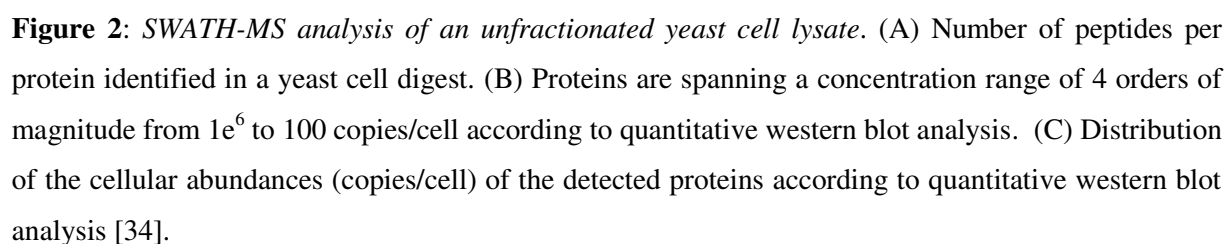


Figure 1: Generation and use of MS/MS spectral libraries and SWATH- bioinformatic tools for the detection and quantification of proteomes and proteins. First, MS coordinates were selected from spectral libraries built from MS/MS spectra acquired by extensive data-dependent analysis experiments of fractionated protein digests. Second, those suitable MS/MS coordinates were used to trigger the fragment ion chromatogram extraction for the peptides of interest in the SWATH maps. Finally, the extracted chromatograms were scored using the probabilistic scoring model of *mProphet* and protein significance analysis was performed by MSstats.



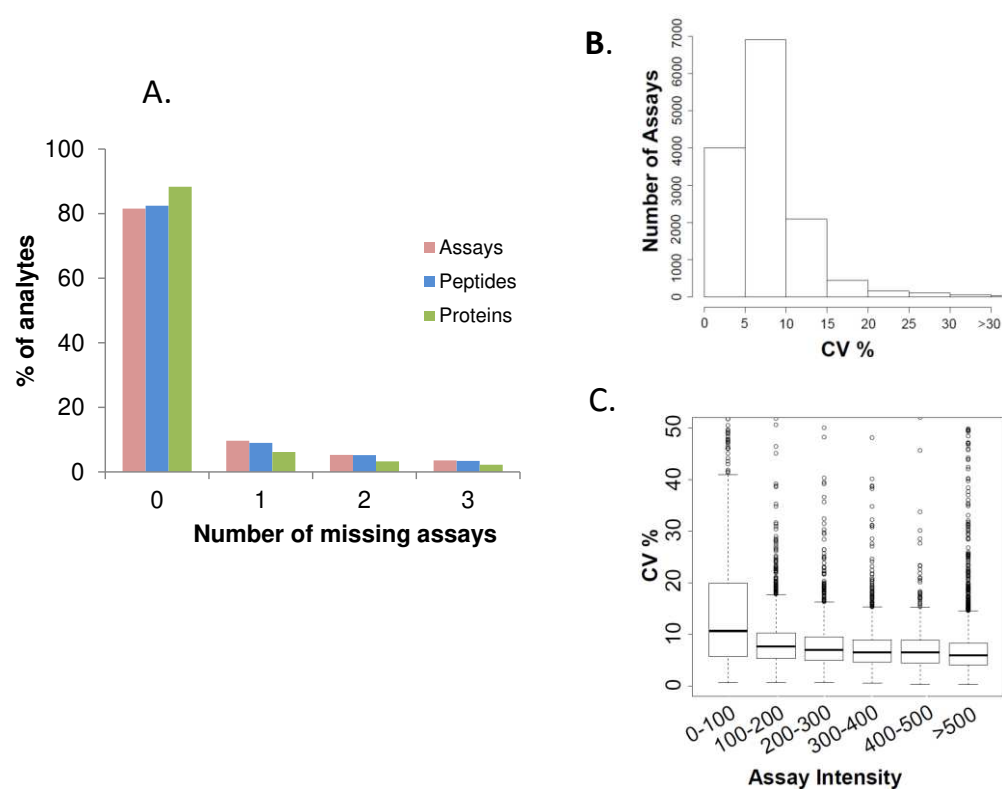


Figure 3: Consistency and reproducibility of SWATH measurements across four injections of an unfractionated tryptic yeast digest. (A) Number of missing assays, peptides and proteins detected across four technical replicates (B) Histogram of CVs of assays detected across all replicates (C) CVs of the assays detected across the entire abundance range and all replicates.

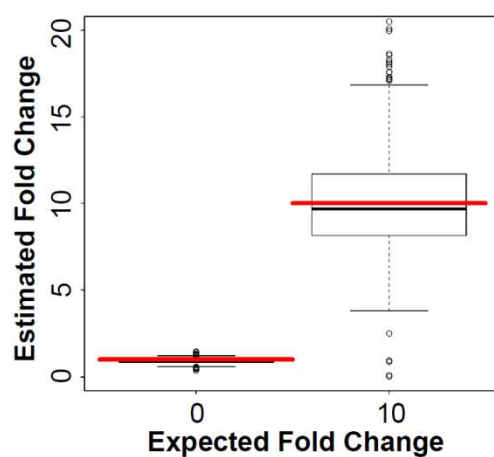


Figure 4: *Accuracy of SWATH measurements.* Estimated fold changes for all the proteins detected in both ^{14}N - and ^{15}N -labeled backgrounds (i.e. 1:1; $^{14}\text{N} : ^{15}\text{N}$ and 1:1/10; $^{14}\text{N} : ^{15}\text{N}$) and all biological replicates. The first box is the estimated fold change of 771 proteins from the study with expected fold change equal to 1 (i.e. 1:1; $^{14}\text{N} : ^{15}\text{N}$). The second box is the estimated fold change of 502 proteins from the study with expected fold change equal to 10 (i.e. 1:1/10; $^{14}\text{N} : ^{15}\text{N}$). The red line in box 1 indicates a fold change of 1. The red line in box 2 indicates a fold change of 10.

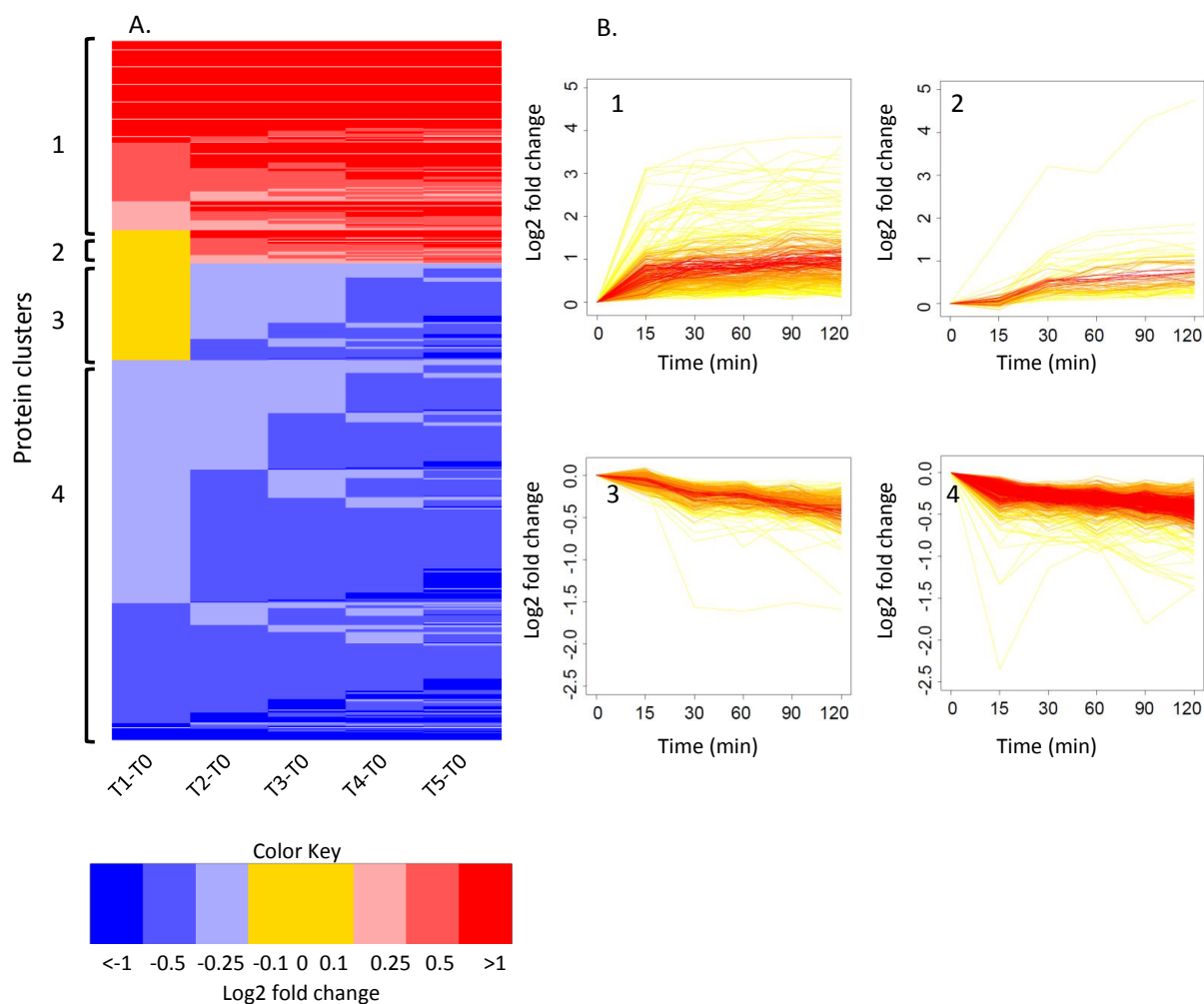


Figure 5: Hierarchical clustering of significantly changing proteins after osmotic shock by SWATH-MS. The osmotic shock was applied by adding 0.4 M NaCl to each 50 ml culture and after 0 min (T0), 15 min (T1), 30 min (T2), 60 min (T3), 90 (T4), 120 min (T5) the cells were harvested and proteins were isolated and digested. (A) In the heatmap, significant up-regulation shown in red, significant down-regulation is shown in blue, and no statistically significant change in abundance is shown in yellow. Color intensity reflects the corresponding log2 fold change. Four main clusters of proteins, according to their patterns of differential abundance in time as compared to the baseline time 0 min (T0) are represented. (B) Expression profile patterns of proteins belonging to the four clusters, where all the time points (i.e T1, T2, T3, T4 and T5) were compared with the time point 0 min (T0).

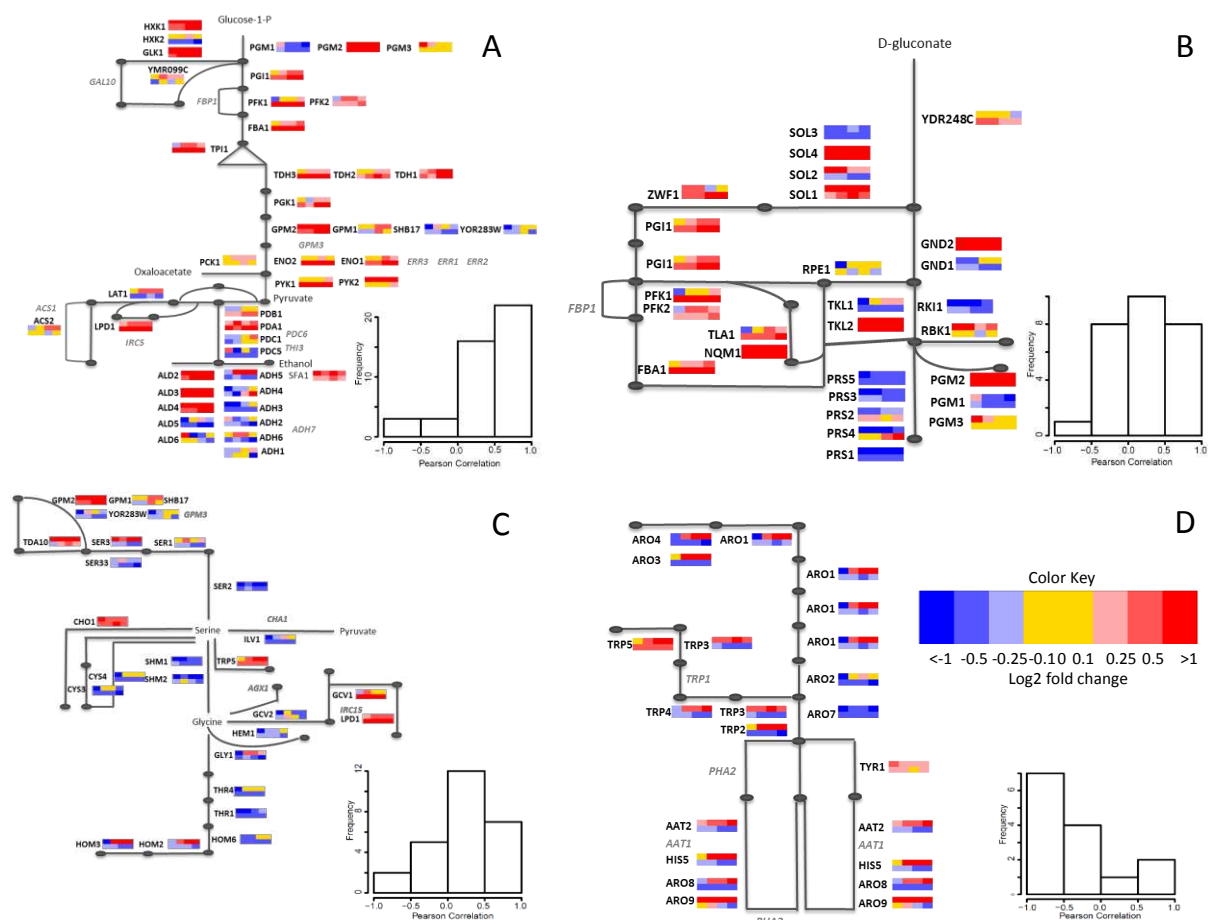


Figure 6: Comparison between proteomics and transcripts data sets. Heat maps for each transcript (top row) and the corresponding protein log₂-fold changes (bottom row) between 30 min (T2), 60 min (T3), 90 min (T4), 120 min (T5) and the initial time point 0 min (T0) together with the distribution of Pearson correlation coefficients calculated for all detected protein-transcript pairs in A. Glycolysis-gluconeogenesis pathway, B. Pentose phosphate pathway, C. Glycine, serine and threonine metabolism pathway and D. Phenylalanine, tyrosine and tryptophan pathway. Significant up-regulation is shown in red, significant down-regulation in blue, and no statistically significant change in abundance is shown in yellow. Color intensity reflects the corresponding log₂ fold change.

Table 1: Distribution of the number of agreements when conducting the test of differential abundance of T1-T0, T2-T0, T3-T0, T4-T0, T5-T0 for SRM and SWATH-MS, respectively. 51 regulated proteins are in our discussion.

Number of Agreements	0	1	2	3	4	5
Protein (%)	3.9	5.9	7.8	17.6	19.6	45.1